Conserved sequences identify the closest living relatives of primates

Mei-Ling Zhang[1,#], Ming-Li Li[2,4,#], Adeola Oluwakemi Ayoola[2,4], Robert W. Murphy[2,3], Dong-Dong Wu[2,*], Yong Shao[2,*]

[1] *Department of Acute Infectious Diseases Control and Prevention*, *Yunnan Provincial Centre for Disease Control and Prevention*, *Kunming Yunnan* 650022, *China*

[2] *State Key Laboratory of Genetic Resources and Evolution*, *Kunming Institute of Zoology*, *Chinese Academy of Sciences*, *Kunming Yunnan* 650223, *China*

[3] *Centre for Biodiversity and Conservation Biology*, *Royal Ontario Museum*, *Toronto M5S* 2C6, *Canada*

[4] *Kunming College of Life Science*, *University of the Chinese Academy of Sciences*, *Kunming Yunnan* 650223, *China*

ABSTRACT

Elucidating the closest living relatives of extant primates is essential for fully understanding important biological processes related to the genomic and phenotypic evolution of primates, especially of humans. However, the phylogenetic placement of these primate relatives remains controversial, with three primary hypotheses currently espoused based on morphological and molecular evidence. In the present study, we used two algorithms to analyze differently partitioned genomic datasets consisting of 45.4 Mb of conserved non-coding elements and 393 kb of concatenated coding sequences to test these hypotheses. We assessed different genomic histories and compared with other molecular studies found solid support for colugos being the closest living relatives of primates. Our phylogeny showed Cercopithecinae to have low levels of nucleotide divergence, especially for Papionini, and gibbons to have a high rate of divergence. The MCMCtree comprehensively updated divergence dates of early evolution of Primatomorpha and Primates.

Open Access

INTRODUCTION

The advancement of genomics has ushered an era in which genetic studies on the evolution of adaptively complicated traits have become possible. This is especially interesting for humans, non-human primates, and their closest relatives. However, such assessment requires a robust hypothesis of phylogenetic relationships. To date, the phylogeny of primates and their closest relatives remains a subject of intense debate. Analyses of different morphological or molecular datasets show conflicting topologies, most likely due to the accelerated evolution of the Euarchonta, which includes the Primates, Southeast Asian flying lemurs (Dermoptera), and tree shrews (Scandentia). Furthermore, apparent incomplete lineage sorting (ILS) can drive conflict between estimations of species trees and gene trees (Murphy et al., 2007; Nishihara et al., 2006; Scornavacca & Galtier, 2017).

Massive parallel data have driven advancements in genomics and associated methods (Liu et al., 2010; Mirarab et al., 2014a, 2014b) and can be applied to hypothesize species trees. Analyses using a greater number of genes or diverse data can be used to parse credible species trees, such as that for modern birds (Jarvis et al., 2014). Longstanding hypotheses based on morphological and molecular proof have suggested a sister-group relationship

between primates and tree shrews (Kay et al., 1992; Novacek, 1992; Wible & Covert, 1987). In addition, several other studies have regarded both tree shrews and flying lemurs (colugos) together as a sister group of primates (Bloch et al., 2007; Liu et al., 2001; Murphy et al., 2001; Nie et al., 2008; Sargis, 2002; Springer et al., 2003, 2004). Further phylogenetic analysis incorporating paleontological evidence also suggests that primates and colugos are sister taxa (Beard, 1993). Previous molecular studies also support colugos as the closest living relatives of primates (Bininda-Emonds et al., 2007; Hudelot et al., 2003; Waddell et al., 2001). Genomic analyses further postulated a third potential topology: ((primates, colugos), tree shrews) (Janecka et al., 2007; Perelman et al., 2011), though this was based on analyses of limited genomic changes (insertion and deletions, InDels) and few nuclear gene fragments. Furthermore, these analyses did not exclude biases due to ILS, data partitioning, or data insufficiency (Chojnowski et al., 2008; Patel et al., 2013; Rokas et al., 2003).

Whole genome data with a well-preserved evolutionary history for the genome of each species can be used to recover an exact species tree (Jarvis et al., 2014; Mitchell et al., 2015; Rokas et al., 2003; Wolf et al., 2002). Herein, we inferred the closest living relatives of primates and estimated the time scale for extant Euarchonta based on 23 existing genome datasets and an out-group (*Mus musculus*). Analyses relied on two types of data: 393 kb of concatenated protein-coding sequences from 706 one-to-one orthologous genes (OOGs) and 45.4 Mb of concatenated conserved non-coding elements (CNEs).

MATERIALS AND METHODS

Identification of OOGs and sequence alignment/trimming analyses

We downloaded all protein-coding sequences of 20 species (primates, colugos, and tree shrews) (Table 1) from the NCBI assembly database (https://www.ncbi.nlm.nih.gov/). Data for *Homo sapiens*, *Pan troglodytes*, *Tupaia belangeri*, and *Mus musculus* were downloaded from Ensembl 88 (http://www.ensembl.org/info/data/ftp/). The dataset with the longest coding sequence (CDS) of each gene for each species was retained and then translated into proteins to detect paired OOGs using InParanoid-4.1 (Remm et al., 2001) (parameters: $score_cutoff=40, $conf_cutoff=0.05, $group_overlap_cutoff= 0.5, $seq_overlap_cutoff=0.5 and $segment_coverage_cutoff= 0.25). Multiple species OOGs were captured by best reciprocal intersections. Prank v100802 (Loytynoja & Goldman, 2005, 2008, 2010) (parameters: −f=fasta −F −codon − noxml − notree − nopost) and Gblocks v0.91b (Castresana, 2000; Talavera & Castresana, 2007) (parameters: −t=c −b4= 5 −b5=n) were used to align orthologous regions and to trim alignments with low-quality regions for each OOG, respectively.

Pairwise and multiple whole genome alignments

The pairwise whole genome alignments of *Pan troglodytes*,

Gorilla gorilla, *Macaca mulatta*, *Microcebus murinus*, *Galeopterus variegatus*, and *Mus musculus* vs. *Homo sapiens* were downloaded from the University of California Santa Cruz (UCSC) pairwise alignments. We also downloaded repeat-masked whole genome sequences (17 species) from the NCBI assembly database, with the repeat-masked *Homo sapiens* genome obtained from the UCSC Genome Browser (http://genome.ucsc.edu/). Pairwise whole genome alignments for 17 species (Table 1) vs. *Homo sapiens* were obtained using LASTZ (Schwartz et al., 2003) with the following parameters: E=30, O=400, K=3 000, L=2 200, and M=50. A 24-way whole genome multiple alignment was then generated using Multiz v11.2 (Blanchette et al., 2004) with default parameters, and with *Homo sapiens* regarded as the reference taxon.

Detection of CNEs and pre-processing

Phastcons (Siepel et al., 2005) was utilized to obtain elements with conserved scores under a given multiple whole genome alignment file and phylo-HMM. The phylo-HMM analysis assumed a conserved and non-conserved state. Modifications of parameters were: --target-coverage 0.3, --expected-length 45, and --rho 0.31. The phylogenetic model for non-conserved regions was produced by phyloFit in the PHAST package (Hubisz et al., 2011) (http://compgen.cshl.edu/phast/). We then used an in-house Python script to extract the CNEs from multiple sequence alignments across multiple genomes. Our screening required that the CNEs fulfilled three criteria: (1) alignments with gaps in at least one of the 24 species were deleted in this study; (2) multiple sequence alignments overlapped with conserved non-coding regions; and (3) length of overlapping sequences exceeded 10 bp.

Estimates of phylogeny, genomic divergence, and divergence time

Codon positions 1, 2, and 1+2 of each filtered OOG among the 24 species were extracted and concatenated. All CNEs were concatenated and trimmed using Gblocks (Castresana, 2000; Talavera & Castresana, 2007). The maximum likelihood (ML) trees using the concatenated genes were reconstructed using RAxML v8.1.17 (Stamatakis, 2014) with the GTR+ GAMMA substitution model and 1 000 bootstraps. We used 1 000 rapid bootstrap replicates to assess branch reliability. Modeltest (Posada & Crandall, 1998) was selected to detect the best substitution model and MrBayes v3.2.6 (Huelsenbeck & Ronquist, 2001) was used to reconstruct a Bayesian inference (BI) tree. The chain length was set to 10 000 000 (10 000 generation/sample), with the first 100 000 samples discarded as burn-in. The reconstructed single-gene trees with a root (*M. musculus*) from RAxML v8.1.17 (Stamatakis, 2014) were applied using maximum pseudo-likelihood estimation of the species tree (MP-EST) (Liu et al., 2010; Shaw et al., 2013) with default parameters to estimate a species tree based on a multispecies coalescent model. This approach was assumed to resolve conflicts between concatenated and coalescent species trees if ILS existed in our dataset. Relative genomic divergence (nucleotide

Table 1 List of taxa, including primates, colugos, tree shrews, and rodents, used in our analyses

Species	Common name	Taxonomy
Homo sapiens	Human	Hominoidea
Pan troglodytes	Chimpanzee	Hominoidea
Gorilla gorilla	Western lowland gorilla	Hominoidea
Pongo abelii	Sumatran orangutan	Hominoidea
Nomascus leucogenys	Northern white-cheeked gibbon	Hominoidea
Macaca mulatta	Rhesus monkey	Old World monkey
Macaca nemestrina	Pig-tailed macaque	Old World monkey
Papio anubis	Olive baboon	Old World monkey
Cercocebus atys	Sooty mangabey	Old World monkey
Mandrillus leucophaeus	Drill	Old World monkey
Chlorocebus sabaeus	Green monkey	Old World monkey
Colobus angolensis	Angolan colobus	Old World monkey
Rhinopithecus roxellana	Golden snub-nosed monkey	Old World monkey
Aotus nancymaae	Ma ˙ s night monkey	New World monkey
Callithrix jacchus	White-tufted-ear marmoset	New World monkey
Cebus capucinus	White-faced sapajou	New World monkey
Saimiri boliviensis	Bolivian squirrel monkey	New World monkey
Carlito syrichta	Philippine tarsier	Tarsiiformes
Propithecus coquereli	Coquerel ˙ s sifaka	Strepsirrhini
Microcebus murinus	Gray mouse lemur	Strepsirrhini
Otolemur garnettii	Small-eared galago	Strepsirrhini
Galeopterus variegatus	Sunda flying lemur	Dermoptera
Tupaia belangeri	Northern tree shrew	Scandentia
Mus musculus	House mouse	Rodentia

substitutions/site) was visualized as branch lengths in the concatenated species phylogeny.

We estimated divergence time using MCMCtree (dos Reis & Yang, 2011; Yang, 2007), with five fossil calibration points used to determine approximate likelihood calculations (Fan et al., 2013; Franzen et al., 2009; Matsui et al., 2009; Poux & Douzery, 2004; Vignaud et al., 2002). Using MCMCtree (dos Reis & Yang, 2011; Yang, 2007), we obtained 10 000 trees with a sampling frequency of 50 and burn-in of the first 10 000 iterations. The parameters were modified as follows: 'clock=3', 'model=0', 'alpha=0', 'ncatG=5', 'cleandata=1', 'BDparas=1 1 0', 'kappa_gamma=6 2', 'alpha_gamma=1 1', 'rgene_gamma= 2 2', 'sigma2_gamma=1 10', and 'finetune=1: .00356 0.02243 0.00633 0.12 .43455'. All other parameters were set to default. Tracer (http://beast.bio.ed.ac.uk/Tracer) was used to detect convergence.

RESULTS AND DISCUSSION

Coding data and phylogeny based on coding partitioned models

We identified 738 OOGs among the 24 species (Table 1). After multiple sequence alignments and removal of ambiguous sites/regions for each gene, 706 OOGs were

retained. Because rapidly evolving nucleotide sites are phylogenetically less informative than slowly evolving ones, especially for ancient groups (Kälersjö et al., 1999), we deleted the third codon positions for each OOG. The remaining data were concatenated and partitioned by codon position as follows: codon position 1 (196 501 bp), position 2 (196 501 bp), and positions 1+2 (393 002 bp). We employed RAxML v8.1.17 (Stamatakis, 2014) to determine phylogeny based on ML algorithm with using the GTR+ GAMMA substitution model. The topologies of the three consensus trees potentially resolved the deep-branch phylogeny of primates. The representative colugos constituted a sister group of primates, which together formed the Primatomorpha. Compared to the concatenated codon 1 and codon 2 sequences, the concatenated codons 1+2 provided greater support for all nodes. Only one node did not have 100% bootstrap support (BS) (Figure 1). The BI analysis was consistent with the sequence data of the concatenated codons 1+2, with high Bayesian posterior probability (BPP) for each node.

Phylogenetic analyses of the three data-partitions were concordant with the ML and BI topologies for Primatomorpha. However, complicated biological processes, such as ILS, can influence the reconstruction of species trees as different
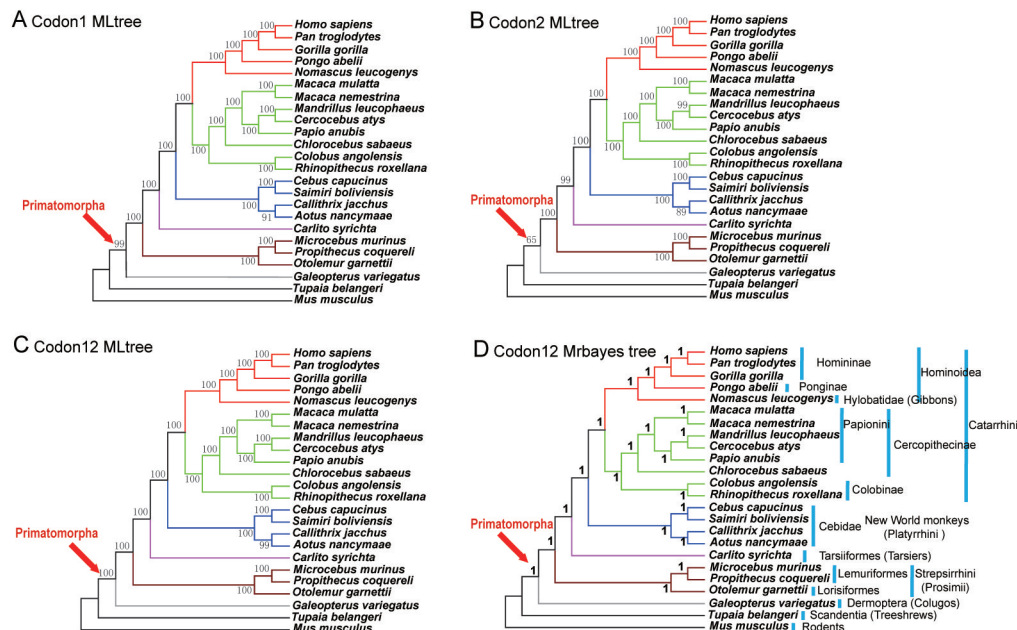
Figure 1 Reconstructed phylogenies based on concatenated sequences from first, second, and first+second codon positions, respectively, of 706 OOGs

(A) and (B) were based on maximum likelihood and (C) and (D) were based on maximum likelihood and Bayesian inference, respectively. Common ancestry branch of primates and colugos was marked by red arrow.

protein-coding genes may have divergent evolutionary histories during the early diversification of Euarchonta (Murphy et al., 2007; Nishihara et al., 2006). Thus, concatenated gene sequences may result in a false-positive species tree with higher nodal BS or BPP (Kubatko & Degnan, 2007). Thus, we estimated the best topology of each gene tree based on the ML algorithm across 24 species, and discovered that discordance between the gene and species trees indeed existed in protein-coding genes in our dataset, which included 100 gene trees (14.1%) supporting tree shrews as the closest living relatives of primates and 25 gene trees (3.5%) supporting both tree shrews and colugos as a sister group of primates. However, we identified a higher proportion of genes (130 genes, ~18.4%) supporting colugos as the closest living relatives of primates (Supplementary Figure S1). Multispecies coalescence can resolve discordance between gene and species trees, even in the presence of ILS (Mirarab et al., 2014a; Song et al., 2012; Zhong et al., 2013). Therefore, we used MP-EST (Liu et al., 2010; Shaw et al., 2013) to construct a coalescent phylogeny. The method maximizes a pseudo-likelihood function to infer a species tree with 1 000 bootstrap replicates. The topology of the coalescent species tree was concordant with the one using the concatenated method. The deep node at the split of colugos and primates had moderate BS (77%) and the node for the last common ancestor of *Callithrix jacchus* and *Aotus nancymaae* had low BS (55.6%) (Supplementary Figure S2). For the phylogeny of primates, the species tree topology based on the concatenated data was consistent with that of

the coalescent data.

Species tree based on conserved non-coding elements

Protein-coding regions only reflect a part of a genome's evolutionary story. Most genome sequences are from non-coding regions. Previous studies have shown that CNEs have important biological functions, including development and transcriptional regulation (Mikkelsen et al., 2007; Woolfe et al., 2005). We used LASTZ (Schwartz et al., 2003) to perform pairwise genome alignments with human as the reference, and then used MULTIZ v11.2 (Blanchette et al., 2004) to obtain whole genome alignments across the 24 species. Phastcons (Siepel et al., 2005) identified 215.6 Mb of concatenated CNEs, with Gblocks v0.91b (Castresana, 2000; Talavera & Castresana, 2007) (parameters: −t=d −b4=5 −b5= n) then used to trim the data for high quality. These analyses obtained 45.4 Mb of data across all 24 species. The RAxML v8.1.17 (Stamatakis, 2014) analyses of the CNEs produced a phylogeny using the GTR+GAMMA substitution model and 1 000 bootstrap replicates. Each of the nodes for the potential species tree arrived at a 100% BS (Figure 2A). This analysis also supported the hypothesis that colugos are the closest living relative of primates. The tree depicted a very short branch length from the last common ancestor of primates and colugos, suggesting that the last common ancestor experienced rapid radiation. This discovery may explain, at least in part, why the sister group of primates is controversial. Analyses using mitochondrial and nuclear sequences did not obtain this result (Murphy et al., 2001; Springer et al., 2003), whereas the globally concatenated data from codons 1+2 plus
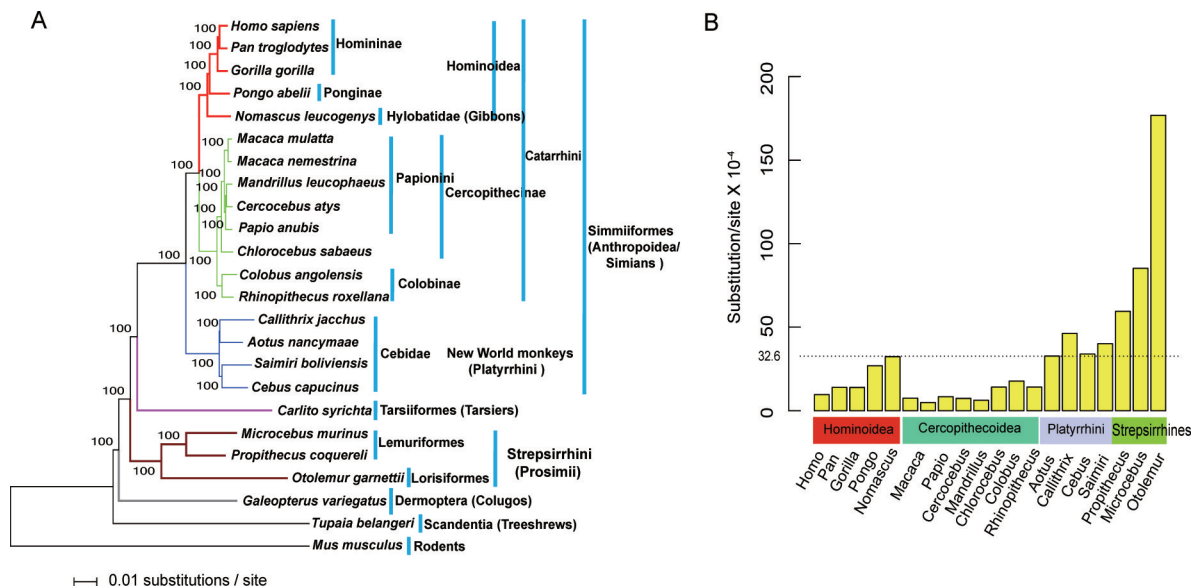
Figure 2 Maximum likelihood tree (ML) and genomic divergences based on conserved non-coding elements (CNEs)
A: CNE ML tree. B: Genomic divergences (nucleotide divergence) of primates. *Carlito syrichta* was excluded because of too few species in Tarsiiformes. Differential groups were marked by blue lines.

CNEs inferred a consistent species tree (Supplementary Figure S3).

CNEs and coding trees vs. InDels and nuclear gene trees

Compared to limited morphological, mitochondrial, and nuclear gene assessments, high-quality whole genomes and robust algorithms can identify, alter, and extend previous taxonomic conclusions. Our phylogenetic analyses, rooted using a rodent (*Mus musculus*), covered the main branches of primates for which genomes exist. Many factors, such as gene duplication, admixture, introgression, hybridization, recombination, convergent evolution, and natural selection, can influence the reconstruction of a true species tree (Degnan & Rosenberg, 2009; Rokas et al., 2003; Scornavacca & Galtier, 2017). However, ILS has often been invoked to explain conflicts between species and gene trees. Many phylogenetic studies examining conflicting signals of different genes have reported considerable discordance across gene trees, e.g., in fruit flies (Pollard et al., 2006) and finches (Jennings & Edwards, 2005). Here, our concatenated and coalescent models supported the hypothesis that colugos are the extant sister group of primates, which together form the Primatomorpha. Concatenated sequences have potentially more informative sites. The coalescent model fills in discordances between the estimated species tree and gene tree once ILS occurs. All methods of analyses in the current study considered different data partitions (protein-coding genes and non-coding sequences), but still obtained a consistent species tree. Our results were in concordance with the hypothesis that colugos are the closest living relatives of primates but refuted the hypotheses that tree shrews or tree shrews plus colugos are the sister group of primates. These observations provide similar insight as two classical molecular studies, which utilized InDels in genomes (Janecka et al., 2007) and 54 sequenced nuclear genes to infer the primate species tree (Perelman et al., 2011). However, our analyses still cannot exclude important influences of many other complex evolutionary histories, such as ancient admixture, introgression, or other potential factors. Recent lines of evidence have indicated that ancient gene flows may potentially underpin phylogenetic discordances (de Manuel et al., 2016; Kuhlwilm et al., 2019; Li et al., 2016; Melo-Ferreira et al., 2014; Nater et al., 2017; Rogers et al., 2019; Tung & Barreiro, 2017; Wu et al., 2018). Therefore, future population genomic studies with deeper coverage and advanced methods need to validate this possibility in clades.

Estimates of genomic divergence in primates

Primate phylogeny can be used to estimate the relative levels of genomic divergences based on branch lengths. The branch uniting Homininae and *Pongo abelii* separated them from *Nomascus leucogenys*, which was assigned to Hylobatidae. This short branch length indicated that Hominoidea radiated rapidly as ((Ponginae, Homininae), Hylobatidae). Overall, primate genomes have varying levels of genomic divergences. According to previous studies, gibbons (Hylobatidae) have an extremely high proportion of chromosomal rearrangement (Jauch et al., 1992; Muller et al., 2003; Perelman et al., 2011). Our analyses indicated that *Nomascus leucogenys* had the highest rate of genomic divergence (32.3 substitutions/site×10^{-4}) among species of Hominoidea. This was almost equivalent to the mean nucleotide divergence (32.6 substitutions/site×10^{-4}) of all primate genomes (Figure 2B). The results of a previous study (Perelman et al., 2011) and our research suggest that this may be a consequence of genomic rearrangements within

Nomascus leucogenys; however, this needs additional investigation. Our results also showed that Old World monkeys had the lowest genomic divergence, and this contrasted with their sister group, Hominoidea and other primates. Low levels of divergence were specific characteristics of Papionini, including *Macaca mulatta*, *Macaca nemestrina*, *Mandrillus leucophaeus*, *Cercocebus atys*, and *Papio anubis*. Speciation patterns of Old World monkeys are remarkably misleading due to convergent morphological traits, behavior, and reproduction, as well as sympatric hybridization (Perelman et al., 2011). Here, the rapid speciation and radiations in Papionini may have involved in reticulate evolution via natural hybridization (Arnold & Meyer, 2006).

Early evolutionary dates of Primatomorpha and Primates

The timing and early evolution of Primatomorpha and Primates are intriguing. A MCMCtree approach (dos Reis & Yang, 2011; Yang, 2007) with five fossil constraints (Fan et al., 2013; Franzen et al., 2009; Matsui et al., 2009; Poux & Douzery, 2004; Vignaud et al., 2002) was used to estimate evolutionary dating (Figure 3). According to previous methods (Liu et al., 2017; Reis et al., 2018), we generated a plot of the posterior distributions to estimate the accuracy of our

divergence dating analyses for calibrated nodes, compared to a previous primate phylogeny study (Perelman et al., 2011) and TimeTree (http://www. timetree. org/) (Supplementary Figure S4). Our analyses were consistent with their results and thus ensured that our analyses were reliable. Our results suggested the origin of Primatomorpha at ~88.0 million years ago (Mya), Primates at ~77.4 Mya, Strepsirrhini at ~70.6 Mya, and Tarsiiformes at ~65.0 Mya (Figure 3). Previously, groups at the base of Euarchonta have been difficult to resolve due to the rapid radiation of ancestral lineages before the Cretaceous-Tertiary (K/T) boundary (~65 Ma) (Wible et al., 2007) and limited sampling of genes and taxa (Janecka et al., 2007; Springer et al., 2003). The estimated dates of origins of Primates, Strepsirrhini, and Tarsiiformes are similar to previous inferences based on nuclear and mitochondrial genes (Jameson et al., 2011; Pozzi et al., 2014). This consistency suggests that our estimates are potentially reliable. These events predated the K/T boundary (Wible et al., 2007). Short branch lengths occurred for the last common ancestor of colugos and primates, as well as for Tarsiiformes and Simiiformes in the CNE tree (Figure 2A). Such rapid divergences may explain why some morphological and molecular studies (Bloch et al., 2007; Sargis, 2002) fail to resolve the placements of colugos and tarsiers. Analyses also
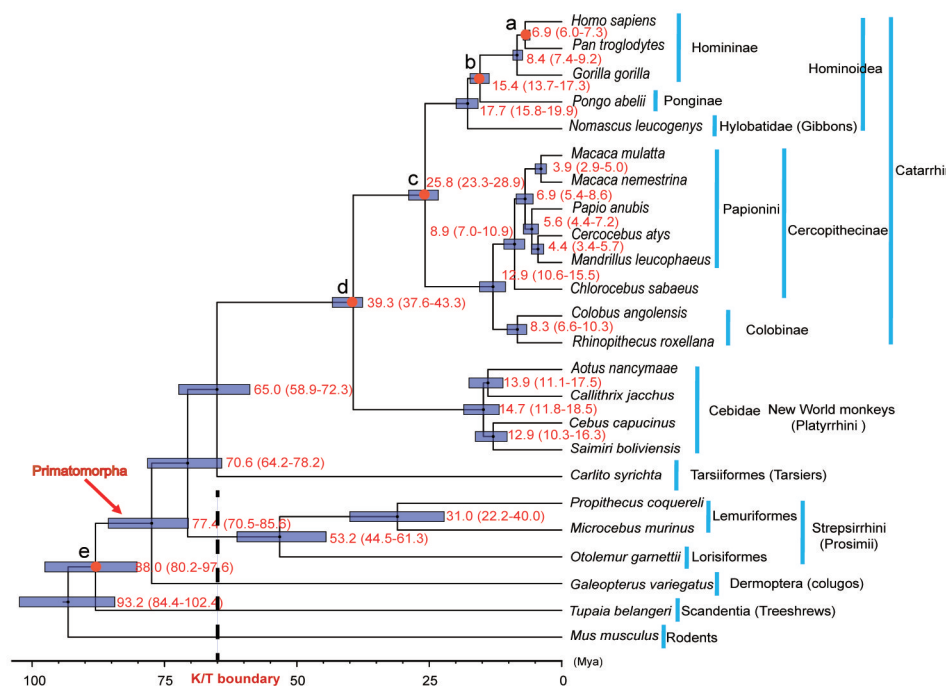
Figure 3 Estimated divergence dates of Euarchonta

Dashed vertical line denoted K/T boundary (~65 Mya). To obtain estimated divergence time, five calibration points with letters a, b, c, d, and e were applied as normal priors to constrain nodal ages. Five red solid points represented fossil constraints, including direct and indirect fossil records: a: mean=6.5 Mya, stdev=0.8 for time to most recent common ancestor (TMRCA) for *Homo-Pan* (Vignaud et al., 2002); b: mean=15.5 Mya, stdev=2.5 for TMRCA of *Homininae-Ponginae* (Matsui et al., 2009); c: mean=29.0 Mya, stdev=6.0 for TMRCA of Catarrhini (Poux & Douzery, 2004); d: mean= 43.0 Mya, stdev=4.5 for TMRCA of Catarrhini-Platyrrhini (Franzen et al., 2009; Poux & Douzery, 2004); e: median=90.9, error range (80.6 – 106.2) for TMRCA of Scandentia-Primatomorpha (Fan et al., 2013). All nodal ages were indicated by medians (red font) and 95% highest posterior density (HPD) intervals (blue bars).

estimated that extant strepsirrhines arose ~53.2 Mya, based on three species (*Propithecus coquereli*, *Microcebus murinus* and *Otolemur garnettii*), which covered two main groups of Strepsirrhini: Lemuriformes and Lorisiformes.

CONCLUSIONS

Our analyses provide a potentially robust assessment of the closest living relatives and divergence times of living primates. The improved conserved non-coding elements and partitioned protein-coding sequences will facilitate further comprehensive studies into the biological processes related to the genomic and phenotypic evolution of primates, especially of humans. Furthermore, our analyses provide a useful resource for the phylogenetic reconstruction of ancient groups.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y. S. and D. D. W. conceptualized the work and revised the paper. M. L. Z. conducted bioinformatic analysis and drafted the paper. M. L. L. analyzed data. A. O. A. and R. W. M. revised the paper. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

REFERENCES

Arnold ML, Meyer A. 2006. Natural hybridization in primates: one evolutionary mechanism. *Zoology*, **109**(4): 261–276.

Beard KC. 1993. Phylogenetic systematics of the Primatomorpha, with special reference to Dermoptera (Chapter 10). *In*: Szalay FS, Novacaek MJ, McKenna MC. Mammal Phylogeny: Placentals. Springer-Verlag, 129–150.

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. *Nature*, **446**(7135): 507–512.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research*, **14**(4): 708–715.

Bloch JI, Silcox MT, Boyer DM, Sargis EJ. 2007. New Paleocene skeletons and the relationship of plesiadapiforms to crown-clade primates. *Proceeding of the National Academy Sciences of the United States of America*, **104**(4): 1159–1164.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**(4): 540–552.

Chojnowski JL, Kimball RT, Braun EL. 2008. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene*, **410**(1): 89–96.

de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, Hernandez-Rodriguez J, Dupanloup I, Lao O, Hallast P, Schmidt JM, Heredia-Genestar JM, Benazzo A, Barbujani G, Peter BM, Kuderna LFK, Casals F, Angedakin S, Arandjelovic M, Boesch C, Kühl H, Vigilant L, Langergraber K, Novembre J, Gut M, Gut I, Navarro A, Carlsen F, Andrés AM, Siegismund HR, Scally A, Excoffier L, Tyler-Smith C, Castellano S, Xue Y, Hvilsom C, Marquew-Bonet T. 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*, **354**(6311): 477–481.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**(6): 332–340.

dos Reis M, Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Molecular Biology and Evolution*, **28**(7): 2161–2172.

Fan Y, Huang ZY, Cao CC, Chen CS, Chen YX, Fan DD, He J, Hou HL, Hu L, Hu XT, Jiang XT, Lai R, Lang YS, Liang B, Liao SG, Mu D, Ma YY, Niu YY, Sun XQ, Xia JQ, Xiao J, Xiong ZQ, Xu L, Yang L, Zhang Y, Zhao W, Zhao XD, Zheng YT, Zhou JM, Zhu YB, Zhang GJ, Wang J, Yao YG. 2013. Genome of the Chinese tree shrew. *Nature Communications*, **4**: 1426.

Franzen JL, Gingerich PD, Habersetzer J, Hurum JH, von Koenigswald W, Smith B. H. 2009. Complete primate skeleton from the Middle Eocene of Messel in Germany: morphology and paleobiology. *PLoS One*, **4**(5): e5723.

Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform*, **12**(1): 41–51.

Hudelot C, Gowri-Shankar V, Jow H, Rattray M, Higgs PG. 2003. RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. *Molecular Phylogenetics and Evolution*, **28**(2): 241–252.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**(8): 754–755.

Jameson NM, Hou ZC, Sterner KN, Weckle A, Goodman M, Steiper ME, Wildman D. E. 2011. Genomic data reject the hypothesis of a prosimian primate clade. *Journal of Human Evolution*, **61**(3): 295–305.

Janecka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ. 2007. Molecular and genomic data identify the closest living relative of primates. *Science*, **318**(5851): 792–794.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, Sun A, Weber CC, da Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan X, Dixon A, Li S, Li N, Huang Y, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello CV, Lovell PV, Wirthlin M, Schneider MPC, Prosdocimi F, Samaniego JA, Velazquez AMV, Alfaro-Núñez A, Campos PF, Perersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong Z, Zeng Y, Liu S, Li Z, Liu B, Wu K, Xiao J, Yinqi X, Zheng Q, Zhang Y, Yang H, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jønsson KA, Johnson W, Koepfli KP, O' Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E, Graves GR, Glenn TC,

McCormack J, Burt D, Ellegren H, Alström P, Edwards SV, Stamatakis A, Mindell DP, Cracraft J, Braun EL, Warnow T, Jun W, Thomas M, Gilbert P, Zhang G. 2014. Whole genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**(6215): 1320–1331.

Jauch A, Wienberg J, Stanyon R, Arnold N, Tofanelli S, Ishida T, Cremer T. 1992. Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proceeding of the National Academy of Sciences of the United States of America*, **89**(18): 8611–8615.

Jennings WB, Edwards SV. 2005. Speciational history of Australian grass finches (Poephila) inferred from thirty gene trees. *Evolution*, **59**(9): 2033–2047.

Kälersjö M, Albert VA, Farris JS. 1999. Homoplasy increases phylogenetic structure. *Cladistics*, **15**(1): 91–93.

Kay RF, Thewissen JGM, Yoder AD. 1992. Cranial anatomy of Ignacius graybullianus and the affinities of the Plesiadapiformes. *American Journal of Physical Anthropology*, **89**(4): 477–498.

Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**(1): 17–24.

Kuhlwilm M, Han S, Sousa VC, Excoffier L, Marques-Bonet T. 2019. Ancient admixture from an extinct ape lineage into bonobos. *Nature Ecology & Evolution*, **3**(6): 957–965.

Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Research*, **26**(1): 1–11.

Liu FG, Miyamoto MM, Freire NP, Ong PQ, Tennant MR, Young TS, Gugel KF. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. *Science*, **291**(5509): 1786–1789.

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology*, **10**: 302.

Liu L, Zhang J, Rheindt FE, Lei F, Qu Y, Wang Y, Zhang Y, Sullivan C, Nie W, Wang J, Chen J, Edwards SV, Meng J, Wu S. 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(35): E7282–E7290.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(30): 10557–10562.

Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, **320**(5883): 1632–1635.

Löytynoja A, Goldman N. 2010. WebPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics*, **11**: 579.

Matsui A, Rakotondraparany F, Munechika I, Hasegawa M, Horai S. 2009. Molecular phylogeny and evolution of prosimians based on complete sequences of mitochondrial DNAs. *Gene*, **441**(1–2): 53–66.

Melo-Ferreira J, Vilela J, Fonseca MM, da Fonseca RR, Boursot P, Alves PC. 2014. The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biology and Evolution*, **6**(4): 886–896.

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome assembly Team, Graves JAM, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh K. 2007. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature*, **447**(7141): 167–177.

Mirarab S, Bayzid MS, Boussau B, Warnow T. 2014a. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science*, **346**(6215): 1250463.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014b. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**(17): i541–i548.

Mitchell KJ, Cooper A, Phillips MJ. 2015. Comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science*, **349**(6255): 1460.

Müller S, Hollatz M, Wienberg J. 2003. Chromosomal phylogeny and evolution of gibbons (Hylobatidae). *Human Genetics*, **113**(6): 493–501.

Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Research*, **17**(4): 413–421.

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stahope MJ, de Jong WW, Springer MS. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, **294**(5550): 2348–2351.

Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M, Desai T, Groves C, Pybus M, Sonay TB, Roos C, Lameira AR, Wich SA, Askew J, Davila-Ross M, Fredriksson G, de Valles G, Casals F, Prado-Martinez J, Goossens B, Verschoor EJ, Warren KS, Singleton I, Marques DA, Pamungkas J, Perwitasari-Farajallah D, Rianti P, Tuuga A, Gut IG, Gut M, Orozco-terWengel P, van Schaik CP, Bertranpetit J, Anisimova M, Scally A, Marques-Bonet T, Meijaard E, Krützen M. 2017. Morphometric, behavioral, and genomic evidence for a new orangutan species. *Current Biology*, **27**(22): 3487–3498.

Nie W, Fu B, O'Brien PCM, Wang J, Su W, Tanomtong A, Volobouev V, Ferguson-Smith MA, Yang F. 2008. Flying lemurs-The 'flying tree shrews'? Molecular cytogenetic evidence for a Scandentia-Dermoptera sister clade. *BMC Biology*, **6**: 18.

Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(26): 9929–9934.

Novacek MJ. 1992. Mammalian phytogeny: shaking the tree. *Nature*, **356**(6365): 121–125.

Patel S, Kimball R, Braun E. 2013. Error in phylogenetic estimation for bushes in the tree of life. *Journal of Phylogenetics & Evolutionary Biology*, **1**(2): 110.

Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MA, Kessing B, Pontius J, Roelke M, Rumpler Y, Schneider MPC, Silva A, O'

Brien SJ, Pecon-Slattery J. 2011. A molecular phylogeny of living primates. *PLoS Genetics*, **7**(3): e1001342.

Pollard DA, Iyer VN, Moses A M, Eisen MB. 2006. Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. *PLoS Genetics*, **2**(10): e173.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**(9): 817–818.

Poux C, Douzery EJ. 2004. Primate phylogeny, evolutionary rate variations, and divergence times: a contribution from the nuclear gene IRBP. *American Journal of Physical Anthropology*, **124**(1): 1–16.

Pozzi L, Hodgson JA, Burrell AS, Sterner KN, Raaum RL, Disotell TR. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **75**: 165–183.

Reis MD, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z, Yoder AD. 2018. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Systematic Biology*, **67**(4): 594–615.

Remm M, Storm CE, Sonnhammer ELL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of Molecular Biology*, **314**(5): 1041–1052.

Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G, Schierup MH, Cheng J, Munch K, Walker JA, Konkel MK, Jordan V, Steely CJ, Beckstrom TO, Bergey C, Burrell A, Schrempf D, Noll A, Kothe M, Kopp GH, Liu Y, Murali S, Billis K, Martin FJ, Muffato M, Cox L, Else J, Disotell T, Muzny DM, Phillips-Conroy J, Aken B, Eichler EE, Marquew-Bonet T, Kosiol C, Batzer MA, Hahn MW, Tung J, Zinner D, Roos C, Jolly CJ, Gibbs RA, Worley KC, Baboon Genome Analysis Consortium. 2019. The comparative genomics and complex population history of *Papio* baboons. *Science Advances*, **5**(1): eaau6947.

Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**(6960): 798–804.

Sargis EJ. 2002. Paleontology: Primate origins nailed. *Science*, **298**(5598): 1564–1565.

Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. *Genome Research*, **13**(1): 103–107.

Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Systematic Biology*, **66**(1): 112–120.

Shaw TI, Ruan Z, Glenn TC, Liu L. 2013. STRAW: species tree analysis web server. *Nucleic Acids Research*, **41**: W238–W241.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, **15**(8): 1034–1050.

Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(37): 14942–14947.

Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(3): 1056–1061.

Springer MS, Stanhope MJ, Madsen O, de Jong WW. 2004. Molecules consolidate the placental mammal tree. *Trends in Ecology & Evolution*, **19**(8): 430–438.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9): 1312–1313.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**(4): 564–577.

Tung J, Barreiro LB. 2017. The contribution of admixture to primate evolution. *Current Opinion in Genetics & Development*, **47**: 61–68.

Vignaud P, Duringer P, Mackaye HT, Likius A, Blondel C, Boisserie JR, De Bonis L, Eisenmann V, Etienne ME, Geraads D, Guy F, Lehmann T, Lihoreau F, Lopez-Martines N, Mourer-Chauviré C, Otero O, Rage JC, Schuster M, Viriot L, Zazzo A, Brunet M. 2002. Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature*, **418**(6894): 152–155.

Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Informatics*, **12**: 141–154.

Wible JR, Covert HH. 1987. Primates: cladistic diagnosis and relationships. *Journal of Human Evolution*, **16**(1): 1–22.

Wible JR, Rougier GW, Novacek MJ, Asher RJ. 2007. Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary. *Nature*, **447**(7147): 1003–1006.

Wolf YI, Rogozin IB, Grishin NV, Koonin EV. 2002. Genome trees and the tree of life. *Trends in Genetics*, **18**(9): 472–479.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJK, Cooke JE, Elgar G. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biology*, **3**(1): e7.

Wu DD, Ding XD, Wang S, Wójcik JM, Zhang Y, Tokarska M, Li Y, Wang MS, Faruque O, Nielsen R, Zhang Q, Zhang YP. 2018. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nature Ecology & Evolution*, **2**(7): 1139–1145.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586–1591.

Zhong B, Liu L, Yan Z, Penny D. 2013. Origin of land plants using the multispecies coalescent model. *Trends in Plant Science*, **18**(9): 492–495.